

# **HICSS-30**

## **Task Force on Network Storage Architecture**

**January 8, 1997**

### **Storage for ASCI**

**Kim Minuzzo (minuzzo1@llnl.gov)**  
**Dave Wiltzius (wiltzius@llnl.gov)**  
**Lawrence Livermore National Lab**

[http://www.llnl.gov/liv\\_comp/siof/siof\\_nap.html](http://www.llnl.gov/liv_comp/siof/siof_nap.html)  
[http://www.llnl.gov/liv\\_comp/siof/hpss\\_nap\\_wg.html](http://www.llnl.gov/liv_comp/siof/hpss_nap_wg.html)



# The ASCI Program

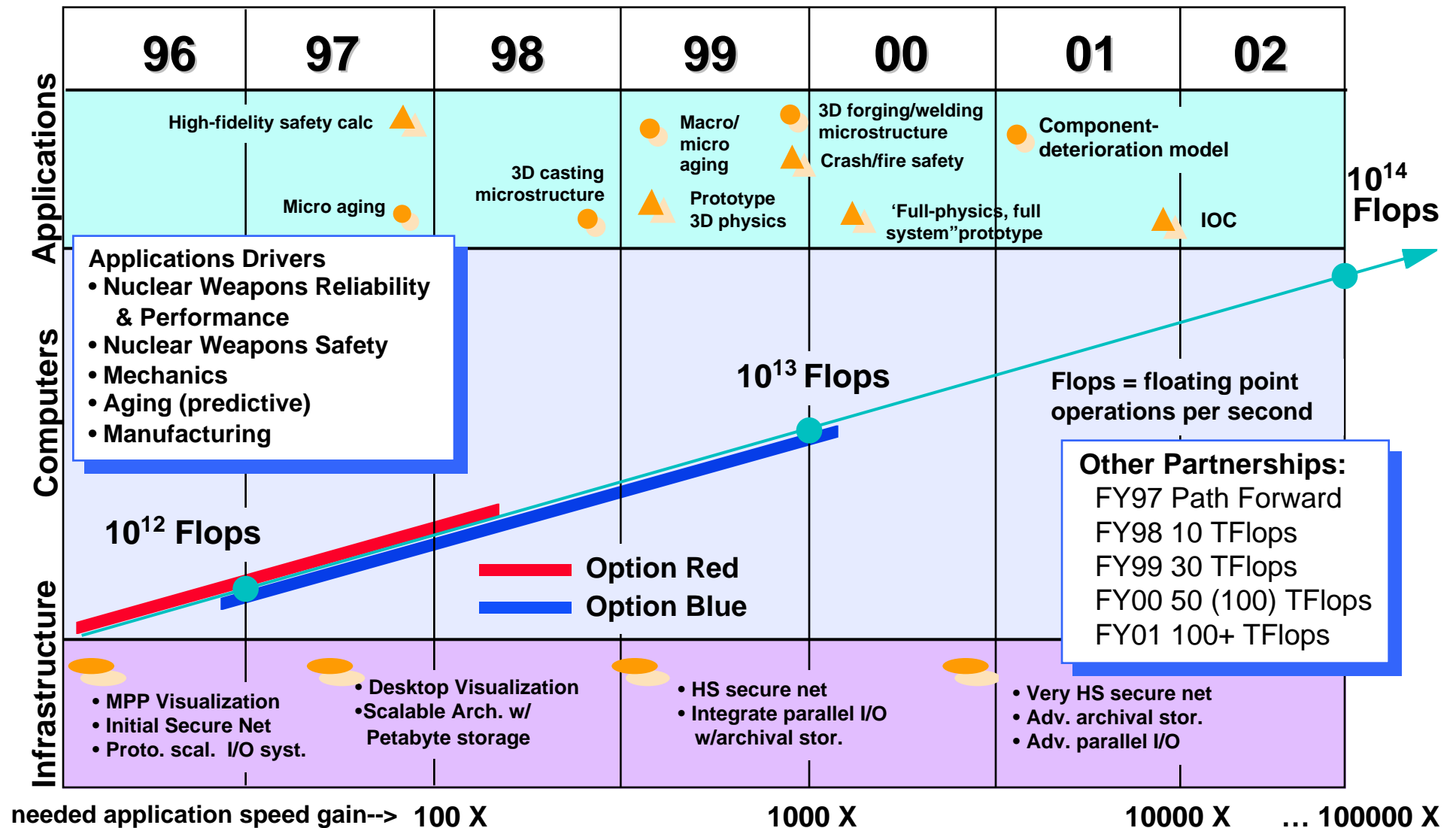


- Accelerated Strategic Computing Initiative (ASCI)
  - » Joint effort of 3 DOE national labs: LLNL, LANL, SNL
  - » 10 year program to 2005
  - » Nuclear stockpile stewardship w/o testing:
    - 3 dimensional, high-resolution nonlinear finite-element applications
  - » Requires super high-performance computational resources
  - » Application driven

<http://www.llnl.gov/asci/>

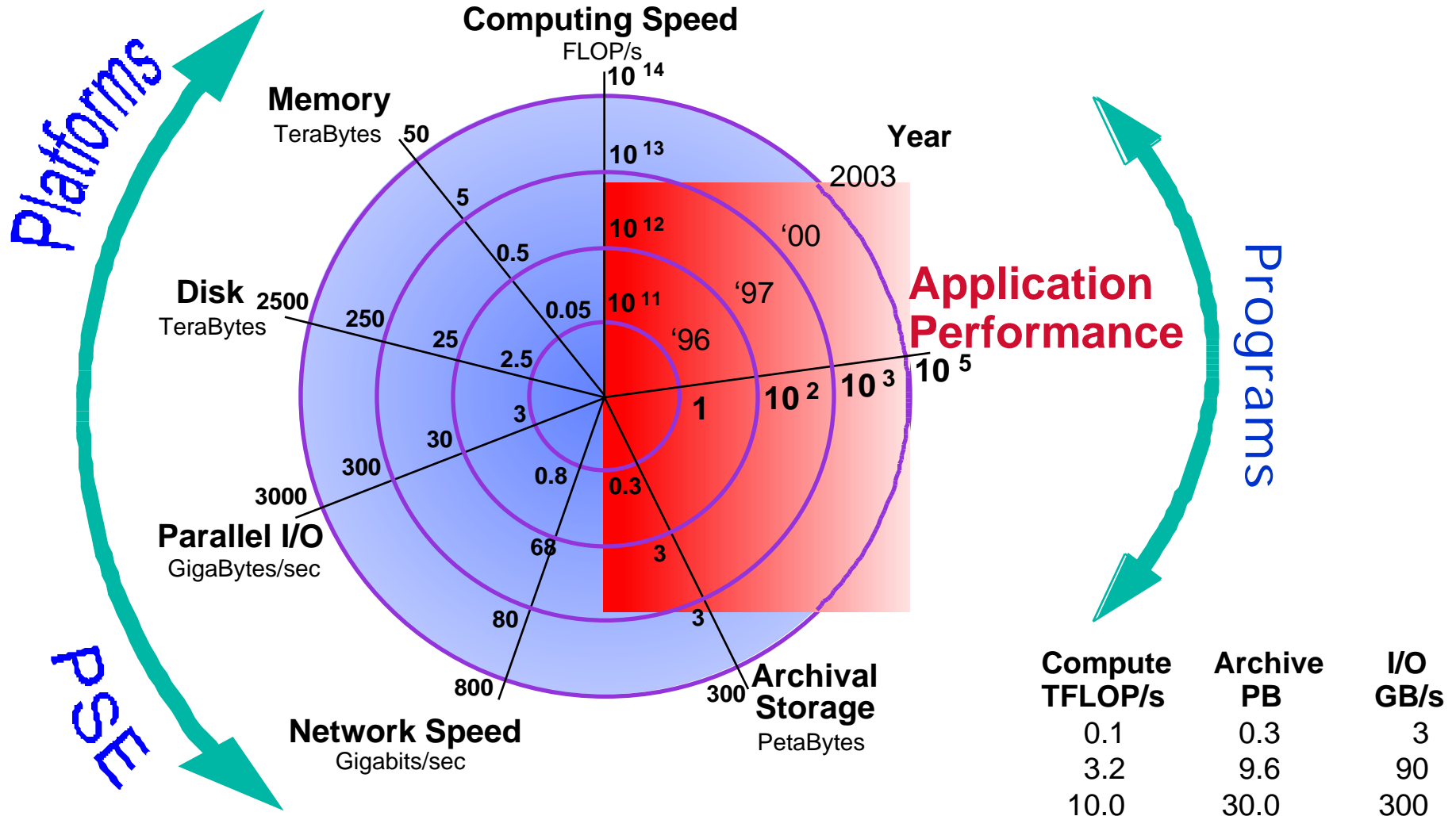


# The ASCI Roadmap





# The key to a usable system is balanced scaling of computational resources





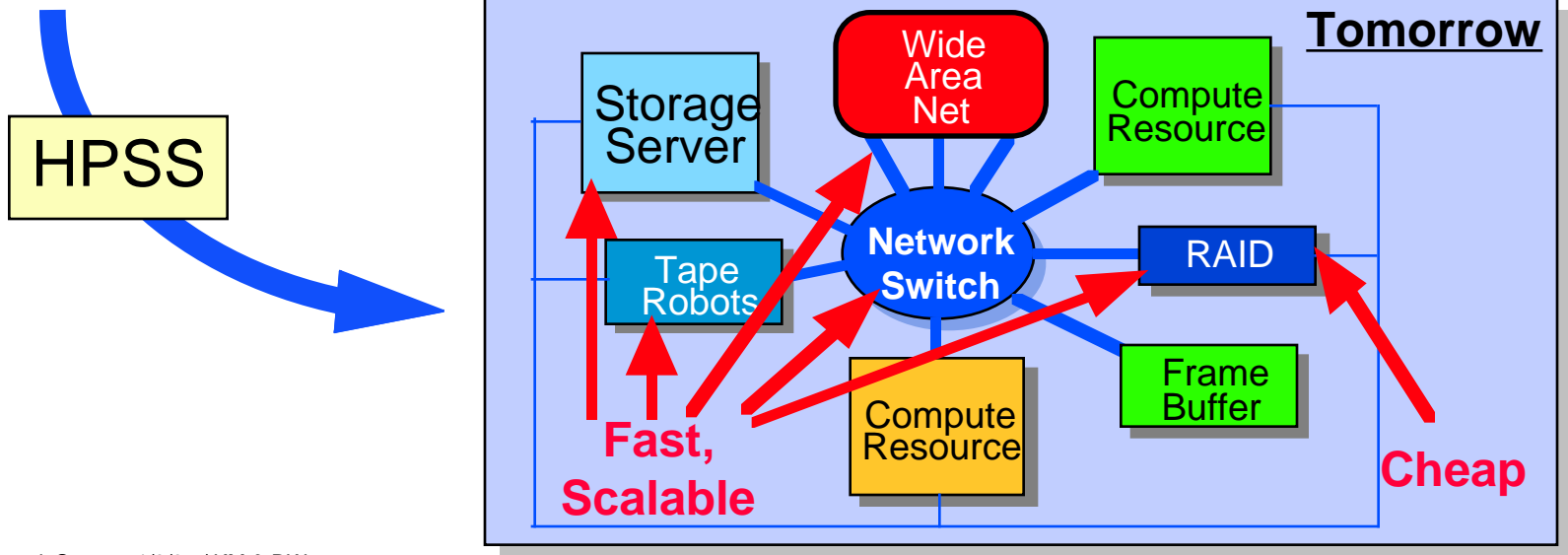
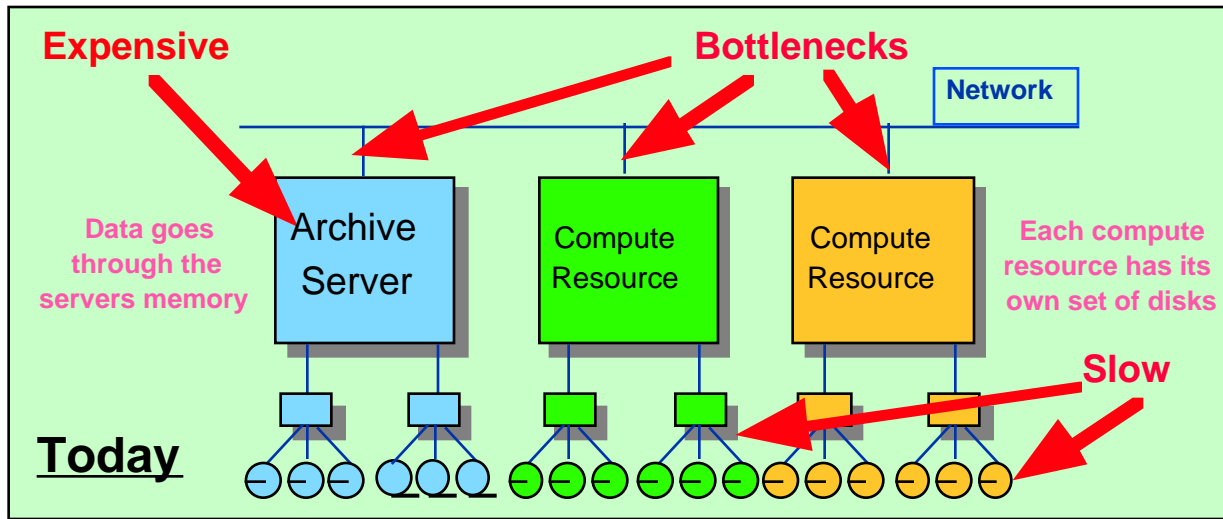
# ASCI Blue - Pacific



- 3.2 TFLOPS/s peak (1 TFLOP/s sustained)
- 2.5 TB of memory
- 75 TB of online disk
- 1 PB of archive (disk and tape) with 10GB/s sustained transfer rate



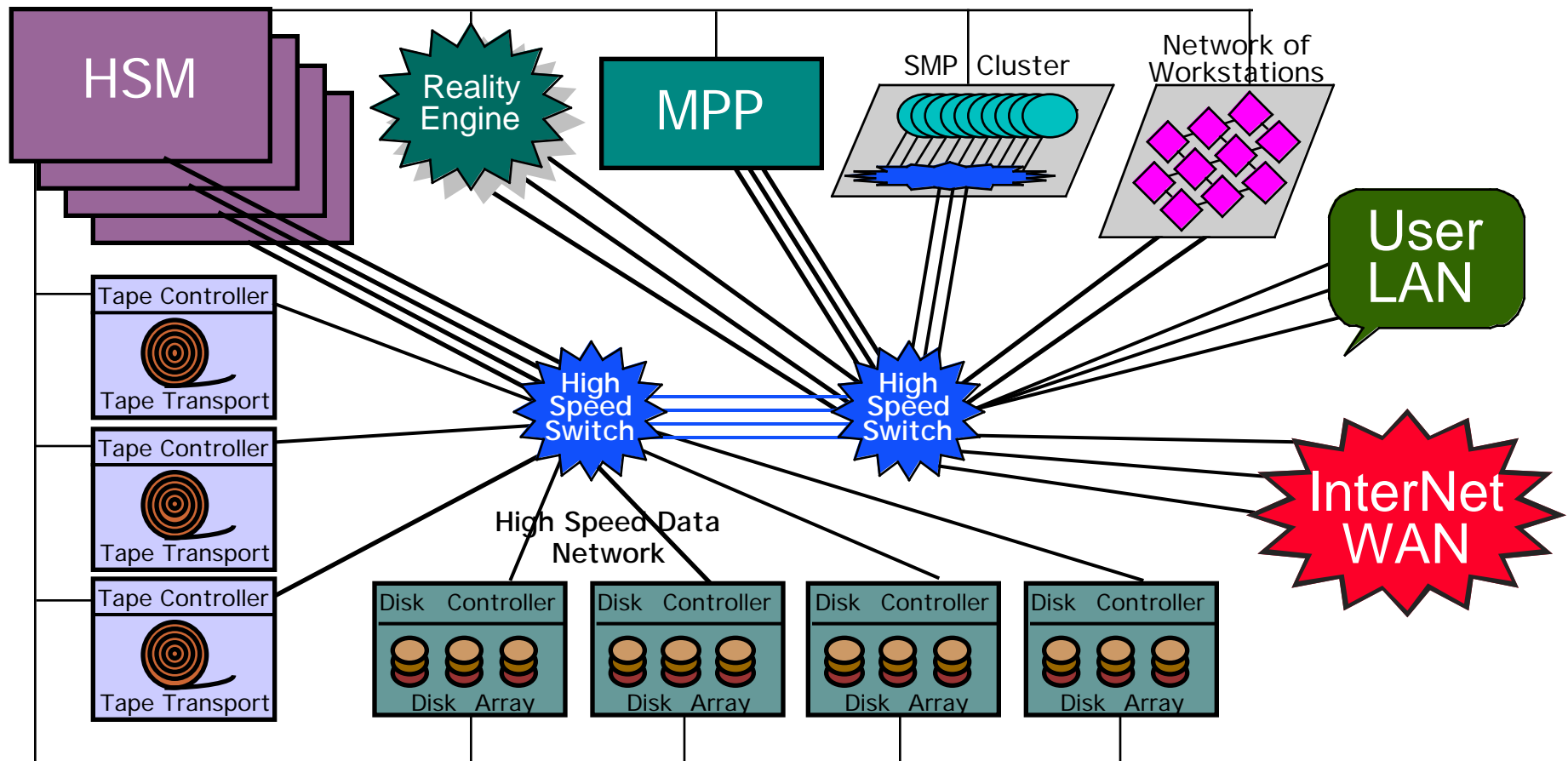
# To achieve performance required by ASCI requires a paradigm shift in HPC storage architecture





# ASCI Architecture

## *The Network Is the Computer!*





# Network storage observations



## ► Data throughput

- » 30-60MB/s throughput at premium price:
  - Customized API (software development/investment)
  - HIPPI hardware (dropping in cost)
  - Customized IPI-3 drivers (high software costs)
  - High-end RAID system (high capital costs)
- » SCSI RAID I/O throughput:
  - 18MB/s via raw I/O, **but**
  - 3MB/s with UNIX I/O

## ► Scalability in capacity and throughput: HSMs can do it!

<u>Number of Clients &amp; Disks</u>	<u>Aggregate Transfer Rate (MB/s)*</u>
16	112.1
32	174.7
64	334.0
128	636.9
240	1353.3

\* Data transfers across IBM SP2 interconnect using TCP/IP





# Network storage observations, cont.



- Data abstraction level
  - » The two ends of the spectrum
    - block I/O e.g., SCSI, IPI-3
    - file I/O e.g., NFS, AFS
  - » Scalable performance implies striping
    - Data space accessible as raw device
    - Ability to separate name (metadata) and data spaces
- Security
  - » Physically separate control network (current HPSS)
    - Simple to implement
    - Cost addition to peripheral
    - Not extensible to WAN environment
  - » Encryption / key-mgmt in drive appears to be best model
- Protocols
  - » TCP/IP must be supported!
  - » Support for multiple protocols desired, e.g. SCSI, IPI-3, ATM .....



# Summary



## Competitively priced network storage strategically important to ASCI!

- network throughput bottlenecks must be alleviated
- data abstraction level - need ability to separate name and data space to allow striping
- security - hardware assisted authentication
- protocols - incorporate adaptive protocol selection - TCP/IP a must for WAN access
- HSMs can successfully utilize network attached storage